

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-240762  
(43)Date of publication of application : 11.09.1998

(51)Int.Cl. G06F 17/30

(21)Application number : 09-046384

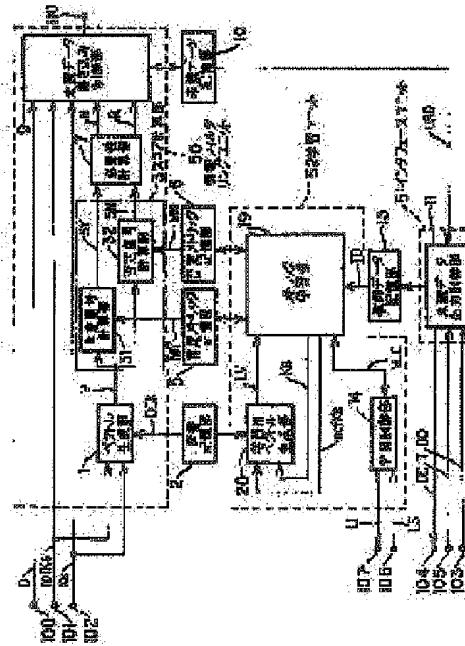
(71)Applicant : MATSUSHITA ELECTRIC IND CO LTD  
(72)Inventor : KANEMICHI TOSHIKI  
YOSHIDA HIDEYUKI  
WATANABE TAISUKE

(54) INFORMATION FILTER DEVICE AND DATA BASE RE-CONSTRUCTING DEVICE AND INFORMATION FILTERING METHOD AND INITIALIZATION METHOD

(57) Abstract:

**PROBLEM TO BE SOLVED:** To provide an information filter device and a method in which information can be provided for a user in the order of information whose necessity is higher by arranging information according to the necessity of the user.

&lt;div



**LEGAL STATUS**

[Date of request for examination] 21.01.1999

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3260093

[Date of registration] 14.12.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of

### rejection]

[Date of extinction of right]

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平10-240762

(43)公開日 平成10年(1998)9月11日

(51)Int.Cl.<sup>6</sup>

G 06 F 17/30

識別記号

F I

G 06 F 15/403

350 C

340 B

審査請求 未請求 請求項の数20 O.L (全 23 頁)

(21)出願番号 特願平9-46384

(71)出願人 000005821

松下電器産業株式会社

大阪府門真市大字門真1006番地

(22)出願日 平成9年(1997)2月28日

(72)発明者 金道敏樹

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

(72)発明者 吉田秀行

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

(72)発明者 渡辺泰助

神奈川県川崎市多摩区東三田3丁目10番1

号 松下技研株式会社内

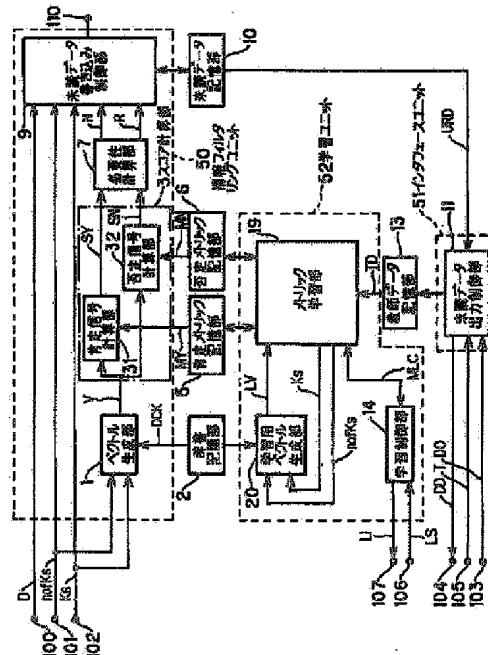
(74)代理人 弁理士 藤合正博

(54)【発明の名称】 情報フィルタ装置とデータベース再構築装置及び情報フィルタリング方法と初期化方法

(57)【要約】

【課題】 情報をユーザーの必要度にしたがって並べ、ユーザーに対して必要性の高い情報から順に提供できる情報フィルタ装置とその方法を提供するものである。

【解決手段】 ベクトル生成部1により情報に割り振られた複数のキーワードをベクトルに変換し、スコア計算部3により前記ベクトルと使用者からの教師信号を用いてスコアを計算し、必要性計算部7により前記スコアから必要性と信頼性を計算し、メトリック学習部19によりスコア計算部3がスコアを計算する際に用いるメトリックを使用者から与えられる情報の必要／不要という単純な評価をもとに計算する。



## 【特許請求の範囲】

【請求項1】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置であって、情報の提示の順序付けのために、複数のキーワード信号を辞書を用いてベクトル信号に変換する手段と、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア信号及び否定スコア信号を計算する手段と、前記肯定スコア信号と前記否定スコア信号との差を必要性信号として計算する手段とを備え、前記必要性信号の大きさにより情報の提示の順序を決めることを特徴とする情報フィルタ装置。

【請求項2】 肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の(i j)成分は、必要とされた情報の頻度と不要とされた情報の頻度と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が必要とされた頻度と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項1記載の情報フィルタ装置。

【請求項3】 行列の(i j)成分は、情報が必要であるか不要であるかを示す確率分布と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項2記載の情報フィルタ装置。

【請求項4】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置であって、情報の提示の順序付けのために、複数のキーワード信号を辞書を用いてベクトル信号に変換する手段を有し、前記ベクトル信号と情報が必要か否かという入力から計算されたベクトルで表現されるメトリック信号からスコア信号を計算し、そのスコア信号を利用して情報の提示順序を決める情報フィルタ装置。

【請求項5】 メトリック信号は、肯定メトリック信号と否定メトリック信号はそれぞれベクトルであり、前記行列のi成分は、必要とされた情報の頻度と不要とされた情報の頻度と、辞書のi番目のキーワード信号が含まれた情報が必要とされた頻度と、辞書のi番目のキーワード信号が含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項4記載の情報フィルタ装置。

【請求項6】 ベクトルのi成分は、情報が必要であるか不要であるかを示す確率分布と、辞書のi番目のキーワード信号が含まれた情報が必要であるか不要であるか

を示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項5記載の情報フィルタ装置。

【請求項7】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置であって、情報の提示の順序付けのために、複数のキーワード信号の辞書を用いてベクトル信号に変換する手段とを有し、前記辞書は、キーワード検索式を含むことを特徴とする情報フィルタ装置。

【請求項8】 辞書は、キーワード検索式を含むことを特徴とする請求項1から6のいずれかに記載の情報フィルタ装置。

【請求項9】 未読の情報を記憶する未読データ記憶部と、前記未読の情報を必要性信号の大きさの順に並べて前記未読データ記憶部に書き込む未読データ書き込み制御部と、前記未読データを順に提示する未読データ出力制御部とを有する請求項1から8のいずれかに記載の情報フィルタ装置。

【請求項10】 情報の必要性を評価するための辞書装置であって、情報が必要か否かを示す信号と前記情報に付けられたひとつまたは複数のキーワード信号とを用いて、情報の必要性を評価するために適するように辞書の内容を変更することを特徴とする適応辞書装置を有することを特徴とする請求項1から9記載の情報フィルタ装置。

【請求項11】 辞書の内容の変更は、必要とされた情報の頻度と、不要とされた情報の頻度と、それぞれのキーワード信号について、前記キーワード信号を含む情報が必要とされた頻度と、前記キーワード信号を含む情報が不要とされた頻度とから計算されるキーワードコスト信号を用いて行われることを特徴とする適応辞書装置を有することを特徴とする請求項10記載の情報フィルタ装置。

【請求項12】 それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする適応辞書装置を有することを特徴とする請求項1から12のいずれかに記載の情報フィルタ装置。

【請求項13】 それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いが大きいほど大きな値の信号であり、前記キーワードコスト信号が大きなキーワード信号を残し、小さいキーワード信号を破棄することを特徴とする適応辞書装置を有することを特徴とする請求項1から12のいずれかに記載の情報フィルタ装置。

【請求項14】 入力端子からの入力が、提示された情報を必要とした回数を示す全肯定回数及び前記情報を不要とした回数を示す全否定回数を記憶する回数記憶部と、キーワード信号を示す文字列を数字に変換する対応表、前記文字列がキーワード信号として含まれた情報を必要とした回数を示す肯定回数及び前記文字列がキーワード信号として含まれた情報を不要とした回数を示す否定回数を記憶した適応辞書記憶部と、提示された情報が必要か否かという入力端子からの入力、前記情報に含まれたキーワード信号、前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号から前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号を更新する辞書学習部とを有することを特徴とする適応辞書装置を有することを特徴とする請求項1から13のいずれかに記載の情報フィルタ装置。

【請求項15】 キーワード信号は、分類コードを含むことを特徴とする請求項1から14のいずれかに記載の情報フィルタ装置。

【請求項16】 情報に対する評価のユーザー入力が必要だけである場合に、情報がユーザーにとって必要である確率を、0と1以外の値とすることを特徴とする請求項3、請求項6、請求項8から15のいずれかに記載の情報フィルタ装置。

【請求項17】 情報に対する評価のユーザー入力が不要だけである場合に、情報がユーザーにとって不要である確率を、0と1以外の値とすることを特徴とする請求項3、請求項6、請求項8から15のいずれかに記載の情報フィルタ装置。

【請求項18】 データベースを再構築する装置であつて、請求項1から17のいずれかに記載の情報フィルタ装置を用いたことを特徴とするデータベース再構築装置。

【請求項19】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ方法であつて、情報の提示の順序付けのために、複数のキーワード信号を辞書を用いてペクトル信号に変換し、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ペクトル信号を用いて肯定スコア信号及び否定スコア信号を計算し、前記肯定スコア信号と前記否定スコア信号との差を必要性信号とし、前記必要性信号の大きさにより情報の提示の順序を決めることを特徴とする情報フィルタリング方法。

【請求項20】 電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、情報が必要であるか不要であるかを示す確率分布と、キーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価することを特徴と

する情報フィルタ装置または情報フィルタ方法の初期化方法であつて、情報がユーザーにとって必要である確率の初期値を、0と1以外の値とする情報フィルタ初期化方法。

#### 【発明の詳細な説明】

##### 【0001】

【発明の属する技術分野】本発明は、電子または光等を媒体とする記憶装置や情報通信網から必要な情報を取り出し易くする情報フィルタ装置とデータベース再構築装置及び情報フィルタリング方法と初期化方法に関するものであり、本願出願人が先に出願した特願平08-230012号の情報フィルタ装置を、より簡単な構成とし、より計算を簡略化し、より好ましい初期設定方法を実現するものである。

##### 【0002】

【従来の技術】近年、情報フィルタ装置は、情報通信の社会基盤の進展に伴い、情報通信網の大規模化と通信量の著しい増大に対応する技術として、その実現が強く望まれている。この背景には、今日、個人が処理可能な情報量に対して、個人がアクセスできる情報量が上回るようになっていることがある。このために、大量の情報の中に必要と思う情報が埋没することが、しばしば起こる。

【0003】情報フィルタ装置に関する従来技術としては、特許検索などに用いられるキーワード論理式をあげることができる。すなわち、数十万から数百万件に及ぶ特許情報をキーワード論理式によりフィルタリングするものである。

##### 【0004】

【発明が解決しようとする課題】しかしながら、キーワード論理式を用いる従来の検索においては、使用者がキーワードについての論理式を精度良く設定する必要があるので、使用者がファイリングされているデータ群の癖（例えば、どのような条件の基に、当該データのキーワードが決定されているのか等）やシステムの構造（例えば、キーワードがシソーラス体系のあるシステムであるか否か等）を十分に知り得ていなければ良い検索ができない。このため、初心者には精度の高い情報フィルタリングを行うことができないという課題があった。

【0005】また、情報フィルタリングした結果もキーワードについての論理式に適合するという評価があるだけであり、たまたまキーワードでは合致しているが、内容は求めているものとは異なるケースであったり、あるいは多くの検索結果から使用者にとって必要度の高い情報をその結果から順に取り出すことは容易ではない。

【0006】本発明は、上記従来の課題を解決するものであり、初心者にも精度の高い情報フィルタリングができ、かつ使用者にとって必要性の高い情報を取り出し易い情報フィルタ装置とデータベース再構築装置及び情報フィルタリング方法と初期化方法を提供することを目的

とする。

【0007】

【課題を解決するための手段】上記目的を達成するためには、本発明の情報フィルタ装置は、情報に割り振られた複数のキーワードをベクトルに変換する手段、前記ベクトルと使用者がどんな情報を必要とし不要としたかを表現した行列を用いてスコアを計算する手段、前記スコアから必要性と信頼性を計算する手段、及び前記必要性の大きい順に情報を並べ替える未読データ書き込み制御手段からなる情報フィルタリングユニットと、必要性の大きい順に情報を提示し、提示した情報が必要か不要かというユーザーの評価を入力できるインタフェースユニットと、ユーザーの評価と前記複数のキーワードとからスコア計算に用いる行列を修正する学習ユニットとを含む構成を有している。

【0008】このような構成によって、複数のキーワードは、距離の定義ができない記号から、使用者の必要度を反映したメトリックを用いて距離を定義できるベクトル表現へと変換され、使用者の必要度を量化することができ、使用者は必要性の高い情報から順に情報を得ることができるようになる。

【0009】

【発明の実施の形態】本発明の請求項1に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置であって、情報の提示の順序付けのために、複数のキーワード信号を辞書を用いてベクトル信号に変換する手段と、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア信号及び否定スコア信号を計算する手段と、前記肯定スコア信号と前記否定スコア信号との差を必要性信号として計算する手段とを備え、前記必要性信号の大きさにより情報の提示の順序を決めることを特徴とする情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0010】本発明の請求項2に記載の発明は、肯定メトリック信号と否定メトリック信号はそれぞれ行列であり、前記行列の(i,j)成分は、必要とされた情報の頻度と不要とされた情報の頻度と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が必要とされた頻度と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項1記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0011】本発明の請求項3に記載の発明は、行列の(i,j)成分は、情報が必要であるか不要であるかを示す確率分布と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項2記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

す確率分布と、辞書のi番目のキーワード信号と辞書のj番目のキーワード信号が同時に含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項2記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0012】本発明の請求項4に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置であって、情報の提示の順序付けのために、複数のキーワード信号を辞書を用いてベクトル信号に変換する手段を有し、前記ベクトル信号と情報が必要か否かという入力から計算されたベクトルで表現されるメトリック信号とからスコア信号を計算し、そのスコア信号を利用して情報の提示順序を決める情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0013】本発明の請求項5に記載の発明は、メトリック信号は、肯定メトリック信号と否定メトリック信号はそれぞれベクトルであり、前記行列のi成分は、必要とされた情報の頻度と、辞書のi番目のキーワード信号が含まれた情報が必要とされた頻度と、辞書のi番目のキーワード信号が含まれた情報が不要とされた頻度とから計算されることを特徴とする請求項4記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0014】本発明の請求項6に記載の発明は、ベクトルのi成分は、情報が必要であるか不要であるかを示す確率分布と、辞書のi番目のキーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする請求項5記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0015】本発明の請求項7に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ装置であって、情報の提示の順序付けのために、複数のキーワード信号の辞書を用いてベクトル信号に変換する手段とを有し、前記辞書は、キーワード検索式を含むことを特徴とする情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0016】本発明の請求項8に記載の発明は、辞書は、キーワード検索式を含むことを特徴とする請求項1から7のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0017】本発明の請求項9に記載の発明は、未読の

情報を記憶する未読データ記憶部と、前記未読の情報を必要性信号の大きさの順に並べて前記未読データ記憶部に書き込む未読データ書き込み制御部と、前記未読データを順に提示する未読データ出力制御部とを有する請求項1から8のいずれかに記載の情報フィルタ装置であり、電子ニュース等のように次々と流入してくる情報から必要とする情報を優先的に提示できるので、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0018】本発明の請求項1.0に記載の発明は、情報の必要性を評価するための辞書装置であって、情報が必要か否かを示す信号と前記情報に付けられたひとつまたは複数のキーワード信号とを用いて、情報の必要性を評価するために適するように辞書の内容を変更することを特徴とする適応辞書装置を有することを特徴とする請求項1から9記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0019】本発明の請求項1.1に記載の発明は、辞書の内容の変更は、必要とされた情報の頻度と、不要とされた情報の頻度と、それぞれのキーワード信号について、前記キーワード信号を含む情報が必要とされた頻度と、前記キーワード信号を含む情報が不要とされた頻度とから計算されるキーワードコスト信号を用いて行われることを特徴とする適応辞書装置を有することを特徴とする請求項1.0載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0020】本発明の請求項1.2に記載の発明は、それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号を含む情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価する信号であることを特徴とする適応辞書装置を有することを特徴とする請求項1から1.2のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0021】本発明の請求項1.3に記載の発明は、それぞれのキーワード信号のキーワードコスト信号は、情報が必要であるか不要であるかを示す確率分布と、前記キーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いが大きいほど大きな値の信号であり、前記キーワードコスト信号が大きなキーワード信号を残し、小さいキーワード信号を破棄することを特徴とする適応辞書装置を有することを特徴とする請求項1から1.2のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0022】本発明の請求項1.4に記載の発明は、入力端子からの入力が、提示された情報を必要とした回数を示す全肯定回数及び前記情報を不要とした回数を示す全

否定回数を記憶する回数記憶部と、キーワード信号を示す文字列を数字に変換する対応表、前記文字列がキーワード信号として含まれた情報を必要とした回数を示す肯定回数及び前記文字列がキーワード信号として含まれた情報を不要とした回数を示す否定回数を記憶した適応辞書記憶部と、提示された情報が必要か否かという入力端子からの入力、前記情報に含まれたキーワード信号、前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号から前記全肯定回数、前記全否定回数及び前記適応辞書記憶部に記憶された信号を更新する辞書学習部とを有することを特徴とする適応辞書装置を有することを特徴とする請求項1から1.3のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0023】本発明の請求項1.5に記載の発明は、キーワード信号は、分類コードを含むことを特徴とする請求項1から1.4のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0024】本発明の請求項1.6に記載の発明は、情報に対する評価のユーザー入力が必要だけである場合に、情報がユーザーにとって必要である確率を、0と1以外の値とすることを特徴とする請求項3、請求項6、請求項8から1.5のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0025】本発明の請求項1.7に記載の発明は、情報に対する評価のユーザー入力が不要だけである場合に、情報がユーザーにとって不要である確率を、0と1以外の値とすることを特徴とする請求項3、請求項6、請求項8から1.5のいずれかに記載の情報フィルタ装置であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0026】本発明の請求項1.8に記載の発明は、データベースを再構築する装置であって、請求項1から1.7のいずれかに記載の情報フィルタ装置を用いたことを特徴とするデータベース再構築装置であり、必要性の高いデータにアクセスしやすいデータベースを再構築できるという作用を持つ。

【0027】本発明の請求項1.9に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、少なくとも入力端子からの入力を用いて情報の提示の順序付けの仕方を変えることを特徴とする情報フィルタ方法であって、情報の提示の順序付けのために、複数のキーワード信号を辞書を用いてベクトル信号に変換し、必要な情報から構成される肯定メトリック信号、不要な情報から構成される否定メトリック信号及び前記ベクトル信号を用いて肯定スコア信号及び否定スコア信号を計算し、前記肯定スコア信号と前記否定スコア信号との差を必要性信号とし、前記必要

性信号の大きさにより情報の提示の順序を決めるこことを特徴とする情報フィルタリング方法であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0028】本発明の請求項20に記載の発明は、電子又は光を媒体とする情報記憶媒体又は情報通信網から所定の情報を取り出すために、情報が必要であるか不要であるかを示す確率分布と、キーワード信号が含まれた情報が必要であるか不要であるかを示す確率分布との違いを定量的に評価することを特徴とする情報フィルタ装置または情報フィルタ方法の初期化方法であって、情報がユーザーにとって必要である確率の初期値を、0と1以外の値とする情報フィルタ初期化方法であり、必要性の高い順に情報をユーザーに提示できるという作用を持つ。

【0029】以下、本発明の実施の形態について、図1から図12を用いて説明する。

(実施の形態1) 以下、本発明の第1の実施の形態について、図面を参照しながら説明する。図1は本発明の第1の実施の形態の情報フィルタ装置の構成を示すブロック図であり、図2はその構成と動作を分かりやすくするために機能単位にまとめたブロック図である。

【0030】まず最初に、図2を用いて本発明の基本概念を説明する。本発明の情報フィルタ装置における基本概念は、ユーザーがどんな「情報」を過去に必要としたかという履歴に関する記録を記憶した複数の記憶部2、5、6と、「情報」のフィルタリングを行う情報フィルタリングユニット50と、その情報フィルタリングユニット50により実際にフィルタリングされた未読の「情報」(ユーザーがまだ読んでいない情報)を蓄積しておく未読データ記憶部10と、ユーザーが当該未読「情報」を可視できるようにしたディスプレー等のインターフェースユニット51と、ユーザーがどんな「情報」を必要としたかという履歴に関する学習を行う学習ユニット52とからなる。

【0031】以下、上記構成の動作について説明する。なお、以下の説明では既にユーザーがどんな「情報」を過去に必要としたかという履歴は学習済みのこととして説明する。また、以下に単に「情報」と称するものは、当該「情報」に対応する1つ以上のキーワードが付されているものとする。そのキーワードとは、当該「情報」を構成する各単語の一部あるいは全体であっても良いし、当該「情報」を代表するために特別に付したものであっても良い。

【0032】まず、情報フィルタリングユニット50に新たな「情報」が入力されると、情報フィルタリングユニット50は、記憶部2、5、6からユーザーがどのような「情報」を過去に必要としたかという記録を読みだし、前記新たな「情報」の必要性を必要性信号として定量的に評価する。

【0033】次に、その評価された新たな「情報」は、

未読データ記憶部10に、必要性信号が大きい順に過去からの未読「情報」を含めて並ぶように前記入力された「情報」を当該順番に書き込む。

【0034】そして、ユーザーが望めば、インターフェースユニット51では、ユーザーに必要性信号の大きい順に前記新たな「情報」を含めた未読「情報」を1つひとつ提示(例えば、ディスプレーに表示)する。その際に、ユーザーに提示された前記新たな「情報」を含めた未読「情報」の1つひとつがユーザーにとって必要か不要かを示す教師信号をユーザーがインターフェースユニット51を介して入力することにより、インターフェースユニット51では、当該教師信号を受け取り、当該「情報」とその教師信号を学習ユニット52に送る。なお、このユーザーによる教師信号の入力は、学習ユニット52の学習能力をより高めるために実施するものであり、学習ユニット52の学習能力(ユーザーがどんな「情報」を過去に必要としたかという履歴の学習能力)が既に十分に高ければ行う必要はない。次に学習ユニット52では、前記提示した「情報」とその教師信号を用いて、記憶部2、5、6の履歴内容を書き換える。

【0035】以上、本発明の情報フィルタ装置は、より高い学習を通じてユーザーに適応し、ユーザーの求める「情報」を優先的に提示することができる。また、当然のことながら、学習を行っていない初期状態では、ユーザーがどんな「情報」を必要としているのか学習ユニット52では分からないので、全ての入力される新たな「情報」をインターフェースユニット51でユーザーが提示を受ける毎に上述したユーザーによる教師信号の入力は必要であるが、隨時実施する学習を通じてやがてユーザーに適応し、ユーザーの求める「情報」を優先的に提示することができる。

【0036】なお、ユーザーの求める「情報」を優先的に提示するとは、より具体的な使用例で述べれば、ある「情報」データベースの母集団Aを特定のキーワードで検索して「情報」の検索集合Bを得たとしても、当該検索集合Bの「情報」の全てがユーザーにとって必要であるとは限らないし、またユーザーにとっては「情報」の全てが必要であってもその必要順位は当然存在することを前提としている。よって、必要から順に不要、あるいはその必要順位に従って、インターフェースユニット51でユーザーに順に提示することを、ユーザーの求める「情報」を優先的に提示することを意味する。

【0037】さて、本発明において重要な点は、いかに必要性信号(ある「情報」が必要であったとの教師信号)を計算するかである。以下に述べる好ましい実施の形態では、必要性信号は概念的に次のような量として計算される。上述した如く、入力された「情報」にキーワードが添付されている場合を考える。一人のユーザーを考えると、そのユーザーが必要としている「情報」に高い頻度または確率で付いているキーワード集合Aと、不

要としている「情報」に高い頻度または確率で付いているキーワード集合Bと、さらにはいずれにもよく付く、または付かないキーワード集合Cとを考えることができます。

【0038】したがって、前記キーワード集合Aに属するキーワードには正の数値を、前記キーワード集合Bに属するキーワードには負の値を、前記キーワード集合Cに属するキーワードには値0をそれぞれ割り振る。そして、新たに入力された「情報」についている1つ以上のキーワードについてそれが、前記キーワード集合A、B、Cのどのキーワードグループに属するかを判定し、前記割り振られた値を積算するように構成する。

【0039】このように構成すれば、前記新たに入力された「情報」に付いていた複数のキーワードを、キーワード集合Aに属するキーワードが数多く含まれた「情報」(ユーザーが必要とする可能性の高い情報)に対しては大きな正の値を示し、キーワード集合Bに属するキーワードが数多く付いている「情報」(ユーザーが不要とする可能性の高い情報)に対しては大きな負の値を示す数値に変換することができる。このようにして、前記数値を用いてユーザーの必要性を予測することができます。

【0040】本発明では、提示した「情報」とその「情報」に関するユーザーの必要/不要の評価とからキーワード(キーワード共起を含む)への値の割り振りを自動的に行い精度の高い必要性信号の計算を実現し、精度高く必要性の高い順に「情報」を並べえることを実現している。そのために、実施の形態1では、「情報」に付けられた複数のキーワードを一つのベクトルに変換し、ユーザーが必要とした場合と不要とした場合について、別々に前記ベクトルの自己相関行列を計算している。ユーザーが必要と答えた「情報」についていたキーワードから作られた自己相関行列MYを用いて、ベクトルVの長さSYを以下の式のように計算する。

【0041】

【数1】

$$DCK[1] = (W[1], C[1])$$

•

•

•

$$DCK[nofDCK] = (W[nofDCK], C[nofDCK])$$

$$SY = \sum_i \sum_j MY_{i,j} \cdot V_i \cdot V_j$$

... (1)

なお、以下の説明では、必要と答えた「情報」についていたキーワードから作られた自己相関行列MYを「肯定メトリック信号」、不要と答えた情報についていたキーワードから作られた自己相関行列MNを「否定メトリック信号」と呼ぶ。この長さSYは、ベクトルVの元となつた複数のキーワードの中に、ユーザーが必要とする「情報」によく含まれているキーワードが数多く含まれていれば、長さSYは大きな正の値をとり、そうでない場合には0に近い値をとるから、必要性信号を計算する上で有効である。

【0042】本発明は、以下に図1を用いて詳細説明するように、さらに工夫を重ねて、精度の高い必要性信号の計算を実現している。図2に示した情報フィルタリングユニット50は、個々の「情報」につけられた複数のキーワード(正確には、分類コードを含む文字列)をベクトルに変換する部分と、ユーザーがどんな「情報」を必要/不要としたという履歴を表現した肯定メトリック信号及び否定メトリック信号を用いてある種のスコアを表す肯定信号と否定信号を計算する部分と、この肯定信号と否定信号とから「情報」の必要性をよく反映する必要性信号を計算する部分と、この必要性信号の大きい順に情報を並べ替える部分からなる。

【0043】以下、情報フィルタリングユニット50に相当するブロックの構成を、図1に即して説明する。図1において、1は「情報」に付けられたキーワードなどの複数の文字列をベクトルに変換するベクトル生成部、2はキーワードなどの複数の文字列をベクトルに変換するための符号辞書信号を記憶した辞書記憶部である。この辞書記憶部2に記憶された符号辞書信号は、「情報」についているキーワードなどの文字列Wを数字Cに変換する対応表をnofDCK個有するコードブック

であり、ベクトル生成部1は、キーワード数信号nofKsとnofKs個のキーワード信号からなるキーワード群信号Ks = (K[1], ..., K[nofKs])とを受けキーワード群信号Ksと前記符号辞書信号DCKを用いてベクトル信号Vに変換する。3はスコア計算部で、ユーザーに提示された「情報」を必要/不要と評価した結果から計算された肯定メトリック信号MY、否定メトリック信号MNを用いて、ベクトル生成部1で変換された2つのベクトル信号Vの長さ、肯定信号SYと否定信号S

... (2)

Nに変換する。5は(nofDCK×nofDCK)行列である前記肯定メトリック信号MYを記憶する肯定メトリック記憶部、6は(nofDCK×nofDCK)行列である前記否定メトリック信号MNを記憶する否定メトリック記憶部である。

7は前記肯定信号SYと前記否定信号SNを受け必要性信号Nと信頼性信号Rを計算する必要性計算部である。

9は「情報」の本文である情報データDとキーワード数信号nofKsとキーワード群信号Ksと必要性信号Nと信頼性信号Rとを所定の手続きに従って後述する未詳デ

タ記憶部10に書き込む未読データ書き込み制御部、10は前記「情報」の本文である情報データDと前記キーワード数信号nofKsと前記キーワード群信号Ksと前記

$$URD\ 1 = (N\ 1, R\ 1, nofKs\ 1, Ks\ 1, D\ 1)$$

•

$$URD\ nofURD = (N\ nofURD, R\ nofURD, nofKs\ nofURD, Ks\ nofURD, D\ nofURD) \dots (3)$$

を記憶する未読データ記憶部、13は最大nofTD個の教師データ信号

$$TD\ 1 = (T\ 1, TnofKs\ 1, TKs\ 1)$$

•

$$TD\ nofTD = (T\ nofTD, TnofKs\ nofTD, TKs\ nofTD)$$

••• (4)

を記憶する教師データ記憶部である。

【0044】次に、図2で示したインタフェースユニット51のブロックの構成を説明する。図1において、11は制御信号DOを受け未読データ記憶部10から未読データ信号URD[1]を読み出し、表示信号DDを出力し、その表示信号DDがユーザーにとって必要か否かを示す教師信号Tをユーザーから受け、前記教師信号Tと前記未読データ信号URD[1]のキーワード数信号nofKs[1]とキーワード群信号Ks[1]とを所定の手続きに従って教師データ記憶部13に書き込む未読データ出力制御部である。

【0045】次に、図2で示した学習ユニット52に相当するブロックの構成を説明する。学習ユニット52は、ユーザーから入力された教師信号Tを用いて肯定／否定メトリック信号を修正するメトリック学習を行う部分と、肯定／否定信号から必要性信号を計算するためのパラメータ、判定パラメータ信号、を修正する部分からなり、各部分は学習制御部14によって制御される。

【0046】図1において、1-9は肯定メトリック記憶部5に記憶された肯定メトリック信号MYと前記否定メトリック記憶部6に記憶された否定メトリック信号MNとを修正するメトリック学習部である。このメトリック学習部1-9は、教師データ記憶部13から前記教師データTDを読み出し、ベクトル生成部1と同じ機能である学習用ベクトル生成部20で複数のキーワードをベクトルに変換し、自己相関行列を計算することで、肯定／否定メトリック信号を修正する。1-4は学習開始信号LSを受けてメトリック学習部1-9を制御する学習制御部である。

【0047】以上のように構成された情報フィルタ装置について、各ユニットごとに図面を用いてその動作を説明する。情報フィルタ装置の好ましい初期状態の一例は、肯定メトリック信号MYと否定メトリック信号MNとを(nofDCK×nofDCK)零行列と、未読データ記憶部1

必要性信号Nと前記信頼性信号Rとからなる最大nofURD個の未読データ

0の未読データURD[i]の全ての必要性信号N[i] (i=1, ..., nofURD)を使用するハードウェアが表現可能な最小の値Vm in、教師データ記憶部20 13の教師データTD[j]の教師信号T[j]を全て-1とした状態である。

【0048】最初に情報フィルタリングユニット50の動作を説明する。まず、情報データ入力端子100から情報データDが入力され、キーワード数信号入力端子101から情報データに付けられたキーワードの個数を表すキーワード数信号nofKsが入力され、キーワード信号入力端子102から複数のキーワードであるキーワード群信号Ks=(K[1], K[2], ..., K[nofKs])が入力される。このベクトル生成部1によってキーワード群信号Ksは、文字列の集まりからベクトル信号Vへと変換される。この変換によって、キーワード群信号の類似性をベクトルの距離として計算できるようになる。

【0049】次に、ベクトル生成部1の動作を図3に示すフローチャートを参照しながら説明する。まず、キーワード数信号nofKsとキーワード群信号Ksを受けると(図3ステップS1)、内部のベクトル信号V=(V[1], V[2], ..., V[nofDic])を(0, 0, ..., 0)に、キーワードカウンタ信号iを1に40セットする(同図ステップS2, S3)。次に、辞書カウンタ信号jを0セットした後辞書カウンタ信号jを1だけ増やす(同図ステップS4)。

【0050】次に、内部のnofDCK個の符号辞書信号DCKを有する辞書記憶部2から辞書カウンタjが指定するキーワードと数字からなる符号辞書信号DCK[j]を読み出し、符号辞書信号DCKの文字列部分W[j]とi番目のキーワード信号K[i]とを比較する(同図ステップS5)。両者が等しくない場合には、辞書カウンタjを1だけ増やす(同図ステップS6)。両者が一致するか、または辞書カウンタjの値が辞書記憶部2に格

納された符号辞書信号の個数nofDicと等しくなるまで図3ステップS5～S7の処理を繰り返す(同図ステップS7)。

【0051】キーワード信号K[i]と等しいW[j]が見つかると、ベクトル信号のj番目の成分V[j]を1にし(同図ステップS8)、キーワードカウンタ信号iを1だけ増やす(同図ステップS9)。以下、同様の処理をキーワードカウンタ信号iがキーワード数信号nofKsより大きくなるまで実行する(同図ステップ(S10))。

【0052】こうして、ベクトル生成部1において、文字列信号からなるキーワード信号の集合体であるキーワード群信号Ks

$$SY = \sum_{i=0}^{nofDic-1} \sum_{j=0}^{nofDic-1} MY[i][j] \cdot V[i] \cdot V[j]$$

... (5)

【0055】否定信号計算部32は、キーワード群信号Ksに過去にユーザーの不要とした情報に含まれていたキーワードが数多く含まれる場合に、大きな値となる否定信号SNを計算する。この目的のために、否定信号計算

$$SN = \sum_{i=0}^{nofDic-1} \sum_{j=0}^{nofDic-1} MN[i][j] \cdot V[i] \cdot V[j]$$

... (6)

【0057】肯定メトリック信号MYと否定メトリック信号MNは、後述するようにキーワード群信号Ksとユーザーの応答に基づいて決められる。本発明では、このように計算された肯定信号SYと否定信号SNを用いて、図9に示したように縦軸に肯定信号SYをとり横軸に否定信号SNをとった2次元空間上の1点に、情報データDを対応させることができる。この2次元空間における情報データDの分布は、ユーザーが必要とするもの(○で表示)は主に左上部に分布し、ユーザーが不要とするもの(×で表示)は主に右下部に分布するようになる。したがって、図8に示したように適切な係数Cを1に設定することにより、ユーザーが必要とする情報データDと不要な情報データDとを分離できる。

【0058】さらに、以下に述べるこの係数Cを用いて計算される必要性信号Nは、上述の2次元空間で左にある程、すなわち、必要性の高いと予測される情報データDほど大きな値となる。したがって、必要性信号Nの大きい順に情報データDを並べて提示すれば、ユーザーは必要な情報を効率よく手に入れることができる。必要性信号Nと直交する方向の信頼性信号Rは、大まかには

ード群信号Ksは、0と1でコード化されたnofDic個の成分を持ったベクトル信号Vに変換される。

【0053】次に、肯定信号計算部31は、キーワード群信号Ksに過去にユーザーの必要とした情報に含まれていたキーワードが数多く含まれる場合に、大きな値となる肯定信号SYを計算する。この目的のために、肯定信号計算部31は、前記ベクトル信号Vを受けて、肯定メトリック記憶部5から肯定メトリック信号MYを読み出し、肯定信号SYを次の式のように計算する。

10 【0054】

【数2】

20 部32は、否定メトリック記憶部6から否定メトリック信号MNを読み出し、否定信号SNを次の式のように計算する。

【0056】

【数3】

30 キーワード群信号Ksに含まれていたキーワードのうちどのくらいのキーワード信号が辞書に含まれていたかを示す信号である。したがって、この信頼性信号Rの大きさは、情報フィルタが計算した必要性信号Nがどれだけ信頼できるのかを示す。

【0059】次に、必要性計算部7は、前記肯定信号計算部31から出力される前記肯定信号SYと前記否定信号計算部32から出力される前記否定信号SNとを受け、過去必要であった情報についていたキーワードが多数あり、不要であった情報についていたキーワードがほとんどない時に大きな値となる必要性信号Nを

$$N = SY - SN$$

と計算し、信頼性信号Rを

$$R = SY + SN$$

と計算する。

【0060】次に、未読データ書き込み制御部9の動作を、図4に示したフローチャートを参照しながら説明する。まず、それぞれの入力端子から前記情報データDと前記キーワード数信号nofKsと前記キーワード群信号Ksと受け、必要性計算部7から前記必要性信号Nと前記信頼性信号Rとを受け、未読データ部指示端子110か

ら出力する未読データ処理信号W1を0から1に変える(図4ステップS11)。次に、i=1とし(同図ステップS12)、未読データ記憶部10に記憶された未読データURD[i]の必要性信号N[i](i=1, ..., n of URD)を順次読み出し、前記必要性信号Nと比較

$$URD[i+1] = URD[i]$$

と置き換え(同図ステップS16～S19)、その後、i1番目の未読データURD[i1]を

$$N[i1] = N$$

$$R[i1] = R$$

$$nofKs[i1] = nofKs$$

$$Ks[i1] = Ks$$

$$D[i1] = D$$

と前記必要性信号N等で置き換える(同図ステップS20)。この置き換えが終了すると、未読データ部指示端子110から出力する未読データ部指示信号W1を0に戻し(同図ステップS20A)、処理を終了する。

【0061】次に、未読データUDRを読みだし、ユーザーの応答(教師信号T)を付加して教師データ信号TDをつくるインターフェースユニット51について説明する。インターフェースユニット51の動作を図5に示したフローチャートを参照しながら説明する。データ読み出し開始信号入力端子103から、データ読み出し開始信号DOが入力される(図5ステップS21)。未読データ出力制御部11は、前記未読データ記憶部10か

$$TD[i] = TD[i-1],$$

と置き換え(同図ステップS28、S29)、1番目の教師データTD[1]を前記教師信号Tと前記未読データのキーワード数信号nofKs[1]とキーワード群信号Ks[1]とを用いて

$$T[1] = T$$

$$TnofKs[1] = nofKs[1]$$

$$TKs[1] = Ks[1]$$

と置き換え(同図ステップS30、S31、S32)、前記未読データ記憶部10の未読データURDをURD[i] = URD[i+1], i = 1, ..., (nofURD-1)とし(同図ステップS33、S34)、nofURD番目の未読データの必要性信号を

$$N[nofURD] = (\text{最小値} V_{min})$$

とする(同図ステップS35、S36、S37)。

【0063】次に、学習ユニット52の動作について図6～図8に示したフローチャートを参照しながら説明する。図6に学習制御部14の動作の概略を示すフローチャート示し、詳しく説明する。図6において、まず、学習開始信号入力端子106から学習開始信号LSが入力され、学習制御部指示信号出力端子107から出力される学習制御部指示信号LIを0から1に変え(図6ステップS41)、処理中を示す。次に、図7のステップS53に対応するメトリック学習部19を動作し(同図ステップS42)、LIを0として(同図ステップS4

較し(同図ステップS13)、前記必要性信号Nが未読データURD[i]の必要性信号N[i]より大きくなる( $N \geq N[i]$ )最初の未読データの番号i1を検出する(同図ステップS14、S15)。そしてi1番目以降の未読データを

$$i = i1, \dots, n of URD$$

ら1番目の未読データURD[1]を読み出し(同図ステップS22)、未読データの必要性信号N[1]が最小値Vminより大きい場合には、未読データ信号URD[1]の情報信号D[1]を表示情報信号DDとしてデータ表示端子104に出力し、待機する(同図ステップS23、S24)。未読データの必要性信号N[1]が最小値Vminに等しい場合には、表示情報信号DDを「データなし」としてデータ表示端子104に出力し、待機する(同図ステップS25)。

【0062】ユーザー(図示せず)は、データ表示装置(図示せず)に表示された表示情報信号DDを見て、それが必要な情報である場合には教師信号T=1、必要でない場合には教師信号T=0、処理を終了する場合には教師信号T=-1として、教師信号入力端子105に返す(同図ステップS26)。教師信号T=-1の場合は処理を終了し、教師信号T≠-1の場合には(同図ステップS27)、未読データ出力制御部11は、教師データ記憶部13の式(4)で表わされる教師データを

$$i = 2, \dots, n of TD$$

3)、処理を終了する。

【0064】次に、メトリック学習部19がユーザーの応答(教師信号T)とキーワード群信号Ksとを用いて、肯定/否定メトリック信号を修正する動作を図7のフローチャートを用いて説明する。図7において、メトリック学習部19は、前記学習制御部14からメトリック学習制御信号MLCを受けると(図7ステップS51)、肯定メトリック記憶部5から肯定メトリック信号MYを、否定メトリック記憶部6から否定メトリック信号MNをそれぞれ読み出し、教師データカウンタcの値を1にする(同図ステップS52)。

【0065】次に、メトリック学習部19は、次に教師データ記憶部13からc番目の教師データ信号TD[c]を読み出し(同図ステップS53)、教師データTD[c]の教師信号T[c]を調べる。前記教師信号T[c]が-1でない場合(T≠-1)には(同図ステップS54)、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する(同図ステップS55)。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述の情報ファイルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する(同図ステップS56)。メトリック学習部

19は、前記学習用ベクトル信号LVを受け、前記教師データTD[c]の教師信号T[c]がT=1である場合には(同図ステップS57)、肯定メトリック信号MYを

$$MY[i][j] = MY[i][j] + LV[i] \cdot LV[j]$$

(ここで、i, j = 1 ~ nofDiC)と修正する(同図ステップS58)。

【0066】この処理により、肯定メトリック信号MYは、ユーザーが必要とした情報データDについていたキーワード信号(複数)に対して大きな値を持つようになる。その結果、前述の肯定信号SYが、ユーザーが必要とする情報データDに対して大きくなるようになる。否定メトリック信号MNも以下のように同様の処理がなされる。

【0067】前記教師データTD[c]の教師信号T[c]がT=0である場合には、否定メトリック信号MNを

$$MN[i][j] = MN[i][j] + LV[i] \cdot LV[j]$$

(ここで、i, j = 1 ~ nofDiC)と修正する(同図ステップS59)。次に、教師データカウンタの値を

$$c = c + 1$$

と1だけ増やす(同図ステップS60)。

【0068】以下、メトリック学習部19は、同様の動作を、教師データTD[c]の教師信号T[c]がT[c]=-1になるかまたはc=nofTDとなるまで繰り返す(同図ステップS61)。T[c]=-1またはc=nofTDとなると(同図ステップS62)、メトリック学習の処理を終了し、メトリック学習制御信号MLCを学習制御部14に送る。学習制御部14は、前記メトリック学習信号MLCを受け、学習制御部指示信号を待機中を示す値にし、処理を終了する。

【0069】図9に示したように、上述の2つのメトリック信号を用いてキーワード群信号を肯定信号SYと否定信号SNとで表される2次元空間上で、ユーザーが必要とする情報は主に左上に、不要な情報は右下に分布するようになる。したがって、上記のように必要性信号をN=SY-SNとすれば、必要性信号は、ユーザーが必要とする情報に対して大きな値をとるようになる。

【0070】また、肯定メトリック信号MYと否定メトリック信号MNの学習を忘却の効果を入れた

$$MY[i][j] = \alpha \cdot MY[i][j] + LV[i] \cdot LV[j]$$

$$MN[i][j] = \beta \cdot MN[i][j] + LV[i] \cdot LV[j]$$

を用いてもよい結果が得られる(ここで、 $\alpha$ と $\beta$ とは、1より小さい正の数)。

【0071】さらに、文献「情報処理学会技術報告、自然言語処理101-8(1994.5.27)」などに

記載された文書からキーワード群信号とキーワード数信号を生成するキーワード生成部を付加する構成をとれば、キーワードが与えられていない情報に対しても適用できる情報フィルタ装置を構成することができる。

【0072】タイトルがつけられた情報については、タイトルを構成する単語をもってキーワードとし、キーワード数信号とキーワード群信号を生成してもよい。加えて、キーワード信号は、国際特許分類番号など分類記号を含むようにしても、本発明の構成を変更する必要はなく、よい結果を得ることができる。

【0073】また、本発明の実施の形態1では、未読データURDを1つずつ提示する場合について示したが、表示装置(図示せず)の大きさによっては、複数の未読データURDを同時に表示し、使用者が複数表示されたどの未読データに対して応答したのかが正しく情報フィルタ装置に伝えられるような構成を取ることは容易である。

【0074】本発明の情報フィルタの根幹は、図7のフローチャートのステップS57、S58、S59に示したように、ユーザーの応答とキーワードとの関係をキーワードの同時出現に注目した肯定メトリック信号MY、否定メトリック信号に反映させ、この2つのメトリック信号を用いてキーワード群信号を肯定信号SYと否定信号SNとに変換することで、キーワードという記号情報を距離の定義された空間に射影したものである。これによって、キーワード群の遠近を距離というアナログ尺度で評価することができるようになる。これを利用することにより、従来の技術では必要か不要かの二者択一的な判定しかできなかった必要性の評価が、ユーザーの必要性の順番に並べるといったことが可能になる。

【0075】以上のように、本発明の実施の形態1の情報フィルタ装置によれば、ユーザーからの教師信号に基づいた学習によって、ユーザーの必要とする情報に対しては必要性信号が大きな値を取るようになり、その結果、表示装置等のインターフェースユニットには、ユーザーにとって必要性が高い情報が優先的に表示されるようになる。

【0076】(実施の形態2)次に、本発明の第2の実施の形態について、図面を参照しながら説明する。本実施の形態2は、実施の形態1の構成に辞書学習部を付加し、辞書記憶部2に記憶された符号辞書信号DCKが使用者に適応するように更新し、かつ肯定メトリック信号MYと否定メトリック信号MNを単純な頻度分布に対応するキーワードの自己相関行列から、情報が必要/不要の出現するキーワードの確率分布を考慮したものへと改良したものである。

【0077】図10に本実施の形態2の情報フィルタ装置のブロック結線図を示すが、以下の説明では、実施の形態1の情報フィルタ装置のブロック結線図と異なる構成についてのみ詳細に説明する。図10において、23

は学習制御部14からの辞書学習信号DLCを受け、辞書記憶部2の符号辞書信号DCKを更新する辞書学習部、24は文字列Wと数字Cがキーワード群信号Ksに含まれていたときに使用者が情報データDを必要と解答した回数を示す肯定回数PYと、文字列Wがキーワード

$$FDCK1 = (W1, C1, PY1, PN1)$$

$$FDCK_{nofFDCK} = (W_{nofFDCK}, C_{nofFDCK}, PY_{nofFDCK}, PN_{nofFDCK}) \quad \dots (7)$$

を記憶した適応符号辞書信号記憶部、25は使用者が必要と答えた回数を示す全肯定回数信号NYと、不要と答えた回数を示す全否定回数信号NNとを記憶する回数記憶部、26は肯定メトリック更新用の1次肯定メトリック信号MY1を記憶する1次肯定メトリック記憶部、27は否定メトリック更新用の1次否定メトリック信号MN1を記憶する1次否定メトリック記憶部、28は前記肯定回数信号と前記否定回数信号と前記1次肯定メトリック信号MY1と前記1次否定メトリック信号MN1とから改良された肯定メトリック信号MYと否定メトリック信号MNを計算してそれぞれを肯定メトリック記憶部5と否定メトリック記憶部6に書き込むKDメトリック学習部である。

【0079】以上のように構成された情報フィルタ装置について、図面を用いて動作を説明する。なお、動作が実施の形態1と同様の箇所は説明を省略する。情報フィルタ装置の好ましい初期状態の一例は、肯定メトリック信号MYと否定メトリック信号MNとを(nofDCK×nofDCK)零行列、未読データ記憶部10の未読データURD

[i]の全ての必要性信号N[i] (i=1, ..., nofURD)を使用するハードウェアが表現可能な最小の値Vmin、教師データ記憶部13の教師データTD

[j]の教師信号T[j]を全て-1、適応符号辞書信号の文字列Wを全てブランク、数字Cを符号辞書信号FDCKの上から順に1、2、..., nofDCK、肯定回数PYと否定回数PNを0、適応符号辞書に対応して、符号辞書の文字列も全てブランクとした状態である。

【0080】まず、情報フィルタリングユニット50は、上述の初期状態の場合、実施の形態1に記載した通りの動作を情報フィルタリングユニット50は行い、入力されたキーワード数信号nofKs、キーワード群信号Ks、情報データDから必要性信号N、信頼性信号Rをともに0と計算し、未読データ記憶部10に格納する。次にインタフェースユニット51は、実施の形態1と同じ動作を行い、使用者の応答が付いた教師データTDを教師データ記憶部13に送る。そして学習ユニット52は、まず学習開始信号入力端子106から学習開始信号LSが入力され、学習制御部14は、前記学習開始信号

群信号Ksに含まれていたときに使用者が情報データDが不要と解答した回数を示す否定回数PNとからなる表をnofFDCK個有する適応符号辞書信号

【0078】

$$FDCK1 = (W1, C1, PY1, PN1)$$

LSを受けて、学習制御部指示信号出力端子107から出力される学習制御部指示信号LIを0から1に変え、処理中を示す。そして辞書学習信号DLCを辞書学習部23に送る。

【0081】以下、図11に示したフローチャートを参照しながら辞書学習部23の動作を説明する。まず、辞書学習信号DLCを受けて(図11ステップS71)、適応符号辞書記憶部24から適応符号辞書FDCKを最大nofDCKtmp個の適応符号信号を記憶できる適応符号信号バッファに読み込み、回数記憶部25から全肯定回数信号NYと全否定回数信号NNとを、1次肯定メトリック記憶部26から1次肯定メトリック信号MY1を、1次否定メトリック信号記憶部27から1次否定メトリック信号MN1を読み出す(同図ステップS72)。次に内部の教師データカウンタcの値を1にし(同図ステップS73)、教師信号記憶部13から教師データTD

[c]を読み出し(同図ステップS74)、その教師信号T[c]が-1であるか否かを判定する(同図ステップS75)。

【0082】T[c] ≠ -1の場合、以下の処理を行う。まず、内部のキーワード数カウンタiの値を1にセットし(同図ステップS76)、適応符号辞書カウンタjの値を1にセットする(同図ステップS77)。次に、前記文字列W[j]がブランクであるかないかを判定し(同図ステップS78)、ブランクである場合には、前記文字列W[j]を前記キーワード信号TK[i]で置き換える(同図ステップS79)。ブランクでない場合には、教師データTD[c]のi番目のキーワード信号TK[i]とj番目の適応符号辞書信号FDCK[j]の文字列W[j]とを比較する(同図ステップS80)。

【0083】前記文字列W[j]がブランクの場合、または、ブランクでなくかつ前記キーワード信号TK[i]と前記文字列W[j]が一致した場合、T[c]の値に応じて以下の処理を行う。T[c] = 1の場合(同図ステップS81)、全肯定信号NYに1を加え(同図ステップS82)、適応符号辞書信号FDCK[j]の肯定回数PY[j]に1を加える(同図ステップS83)。T[c] ≠ 1、これはT[c] = 0の場合

であるが、全否定信号NNに1を加え（同図ステップS84）、適応符号辞書信号FDCK[j]の否定回数PN[j]に1を加える（同図ステップS85）。

【0084】前記W[j]がブランクでなくかつ前記キーワード信号TK[i]と前記文字列W[j]が一致しない場合、適応符号辞書カウンタjの値を1増やす（同図ステップS86）。適応符号辞書カウンタjの値が適応符号辞書信号バッファに記憶できる適応符号信号の数に1を加えた値nofDCKtmp+1と比較する（同図ステップS87）。適応符号辞書カウンタjの値が、nofDCKtmp+1以下の場合、文字列W[j]がブランクかどうかの判定に戻る。それ以外の場合は、前記キーワードカウンタiの値を1だけ増やす（同図ステップS88）。

【0085】前記キーワードカウンタiの値が、前記教師データTD[c]のキーワード数信号TnofKSに1を加えた値TnofKS+1と比較して小さい場合（同図ステップS89）、辞書カウンタjを1にセットし、同様の処理を行う。それ以外の場合、教師データカウンタcの値を1だけ増やす（同図ステップS90）。教師データカウンタcの値と、教師データ数nofIDに1を加えた値nofID+1とを比較し（同図ステップS91）、教師データカウンタcの値が小さい場合、次の教師データTD[c]を読み出して同様の処理を行う。以上の処理が、全ての教師データTDに対して行われる。

【0086】次に辞書学習部23は、各々の適応符号辞

$$\begin{aligned} & NY / (NY + NN) \cdot \log((PY[j] / PY[j] + PN[j])) \\ & + NN / (NY + NN) \cdot \log((PN[j] / PY[j] + PN[j])) \\ & \dots \quad (8) \end{aligned}$$

が考えられる。しかし、これは、そのままでは、本情報フィルタ装置の初期状態など、全肯定回数信号NY、全否定回数信号NN、肯定回数PY[j]、否定回数PN[j]が0のときには、 $\log( )$ の計算ができない、 $PY[j] + PN[j] \approx 1$

$$\begin{aligned} KD[j] = & \tanh((PY[j] + PN[j] / PC) \cdot \tanh(NY / (NY + NN) \cdot \log((PY[j] + \epsilon) / (PY[j] + PN[j] + 2\epsilon)) + NN / (NY + NN) \cdot \log((PN[j] + \epsilon) / (PY[j] + PN[j] + 2\epsilon)))) \\ & \dots \quad (9) \end{aligned}$$

とするものである。ここで、 $\epsilon$ は0でのわり算、 $\log 0$ を避けるための小さな正の値を持つパラメータである。パラメータPCは、3程度の値とするとよい。

【0089】次に、適応符号辞書信号FDCK[j]の文字列W[j]と肯定回数PY[j]と否定回数PN[j]とをキーワードコスト信号KDの大きい順に並べ

$$M[i][j] = MY1[C[i]][C[j]], \quad i, j = 1, \dots, nofDCK$$

その他の場合は、 $i = j$ の場合は、

$$M[i][i] = PY[C[i]], \quad i, = 1, \dots, nofDCK$$

$i \neq j$ の場合は、

$$M[i][j] = 0, \quad i, j = 1, \dots, nofDCK$$

$$MY1[i][j] = M[i][j], \quad i, j = 1, \dots, nofDCK$$

書信号FDCK[j]に対してキーワードコスト信号KDを計算する。このキーワードコスト信号は、文字列W[j]がキーワードとして有効であるか否かを判断するために用いられる量である。

【0087】ところで、使用者の不要な情報データDが出現する確率

$$NN / (NY + NN)$$

と比較して、文字列W[j]が付いている情報データDが使用者にとって不要である場合の確率

$$10 \quad PN[j] / (PY[j] + PN[j])$$

が大きく異なる場合に、大きくなるようものであれば、文字列W[j]は、情報データDが使用者にとって不要と判定する上で有効である。同様に、使用者の必要な情報データDが出現する確率

$$NY / (NY + NN)$$

と比較して、文字列W[j]が付いている情報データDが使用者にとって必要である場合の確率

$$PY[j] / (PY[j] + PN[j])$$

が大きく異なる場合に、大きくなるようものであれば、  
20 文字列W[j]は、情報データDが使用者にとって必要と判定する上で有効である。

【0088】キーワードコスト信号KDは、この性質を反映している量で有ればなんでもよいが、好ましい例の一つとして、カルバックダイバージェンスと呼ばれる

30 を満たす適応符号辞書信号FDCK[j]のキーワードコスト信号を過大評価する等不適切な場合がある。これを回避する好ましい実施の形態の一つは、キーワードコスト信号を

40 替える（同図ステップS92）。このとき、適応符号辞書FDCK[j]の数字C[j]には、最初の並び順が残っている。これをを利用して、1次肯定メトリック信号MY1とC[j]から、C[i]、C[j]の値がともに符号辞書DCKの数nofDCKより小さい場合、

替える（同図ステップS92）。このとき、適応符号辞

書FDCK[j]の数字C[j]には、最初の並び順が

残っている。これをを利用して、1次肯定メトリック信号

MY1とC[j]から、C[i]、C[j]の値がともに符号辞書DCKの数nofDCKより小さい場合、

替える（同図ステップS92）。このとき、適応符号辞

書FDCK[j]の数字C[j]には、最初の並び順が

残っている。これをを利用して、1次肯定メトリック信号

MY1とC[j]から、C[i]、C[j]の値がともに符号辞書DCKの数nofDCKより小さい場合、

替える（同図ステップS92）。このとき、適応符号辞

書FDCK[j]の数字C[j]には、最初の並び順が

残っている。これをを利用して、1次肯定メトリック信号

MY1とC[j]から、C[i]、C[j]の値がともに符号辞書DCKの数nofDCKより小さい場合、

と1次肯定メトリック信号MY1の置き換えを行う。1次否定メトリック信号MN1に対しても、同様の置き換えを行う(同図ステップS93)。そして、適応符号辞書信号バッファ内の適応符号辞書FDCK[j]の数字C[j]を  
 $C[j] = j, j = 1, \dots, \text{nofFCKtmp}$   
 と置き換える。

【0090】以上の処理を終えると、辞書学習部23は、適応符号辞書バッファ内の適応符号辞書FDCKの上位nofDCK個の文字列W[j]と数字C[j]を辞書記憶部2に書き込み、適応符号辞書バッファ内の適応符号辞書信号FDCK[j]の上位nofDCK個を適応符号辞書記憶部24に書き込み、全肯定回数信号NYと全否定回数信号NNを回数記憶部25に書き込み、1次肯定メトリック信号MY1を1次肯定メトリック信号記憶部26に1次否定メトリック信号MN1を1次否定メトリック信号記憶部27に書き込む(同図ステップS94)。最後に、辞書学習信号DCLを学習制御部14に戻し(同図ステップS95)、処理を終了する。

【0091】次に、前記学習制御部14は、KDメトリック学習部28にメトリック学習制御信号MLCを送る。前記メトリック学習制御信号MLCを受けたKDメトリック学習部28は、まず、1次肯定メトリック記憶部26から1次肯定メトリック信号MY1を、1次否定メトリック記憶部27から1次否定メトリック信号MN1をそれぞれ読み出す。

【0092】次に、KDメトリック学習部28は、教師データカウンタcの値を1にする。教師データ記憶部3からc番目の教師データ信号TD[c]を読み出し、教師データTD[c]の教師信号T[c]を調べる。前記教師信号T[c]が-1でない場合(T≠-1)には、教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群信号TKs[c]とを出力する。前記教師データTD[c]のキーワード数信号TnofKs[c]とキーワード群TKs[c]とを受けた学習用ベクトル生成部20は、前述した実施の形態1の情報フィルタリングユニット50のベクトル生成部1と同様の動作を行い、学習用ベクトル信号LVを出力する。KDメトリック学習部28は、前記学習用ベクトル信号LVを受け、前記教師データTD[c]の教師信号T[c]が

$$\text{MY1}[i][j] = \text{NY} / (\text{NY} + \text{NN}) \cdot \log((\text{MY1}[i][j] + \varepsilon) \cdot (\text{NY} + \text{NN}) / (\text{NY} \cdot (\text{MY1}[i][j] + \text{MN1}[i][j] + 2\varepsilon))) \quad \dots (10)$$

と計算し、否定メトリック信号MNを

$$\text{MN1}[i][j] = \text{NN} / (\text{NY} + \text{NN}) \cdot \log((\text{MN1}[i][j] + \varepsilon) \cdot (\text{NY} + \text{NN}) / (\text{NN} \cdot (\text{MY1}[i][j] + \text{MN1}[i][j] + 2\varepsilon))) \quad \dots (11)$$

と計算する。ここで、 $\varepsilon$ は0でのわり算、log0を避けるための小さな正の値を持つパラメータである。

T=1である場合には、1次肯定メトリック信号MY1を

$$\text{MY1}[i][j] = \text{MY1}[i][j] + \text{LV}[i] \cdot \text{LV}[j]$$

(ここで、 $i, j = 1 \sim \text{nofDiC}$ )と修正する。前記教師データTD[c]の教師信号T[c]がT=0である場合には、1次否定メトリック信号MN1を

$$\text{MN1}[i][j] = \text{MN1}[i][j] + \text{LV}[i] \cdot \text{LV}[j]$$

10 (ここで、 $i, j = 1 \sim \text{nofDiC}$ )と修正する。教師データカウンタの値を

$$c = c + 1$$

と1だけ増やす。

【0093】以下、KDメトリック学習部28は、同様の動作を、教師データTD[c]の教師信号T[c]がT[c]=-1になるかまたはc=nofTDとなるまで繰り返す。T[c]=-1またはc=nofTDとなると、1次肯定メトリック信号MY1と1次否定メトリック信号MN1の学習を終える。

20 【0094】次に、回数記憶部25から全肯定回数信号NYと全否定回数信号NNを読み出し、1次肯定メトリック信号MY1と1次否定メトリック信号MN1とを用いて肯定メトリック信号MYを計算する。こうして計算される肯定メトリック信号MY、否定メトリック信号MNは、キーワードコスト信号KDと同様、計算される肯定信号SYと否定信号SNが、使用者の不要な情報データDが出現する確率

$$\text{NN} / (\text{NY} + \text{NN})$$

と比較して、文字列W[j]が付いている情報データDが使用者にとって不要である場合の確率

$$\text{PN}[j] / (\text{PY}[j] + \text{PN}[j])$$

が大きく異なる場合に、大きくなるようなものであり、使用者の必要な情報データDが出現する確率

$$\text{NY} / (\text{NY} + \text{NN})$$

と比較して、文字列W[j]が付いている情報データDが使用者にとって必要である場合の確率

$$\text{PY}[j] / (\text{PY}[j] + \text{PN}[j])$$

が大きく異なる場合に、大きくなるようなものであるといつた性質を持っていれば、なんでもよい。これを満たす好ましいのは、肯定メトリック信号MYを

$$\text{MY1}[i][j] = \text{NY} / (\text{NY} + \text{NN}) \cdot \log((\text{MY1}[i][j] + \varepsilon) \cdot (\text{NY} + \text{NN}) / (\text{NY} \cdot (\text{MY1}[i][j] + \text{MN1}[i][j] + 2\varepsilon))) \quad \dots (10)$$

と計算し、否定メトリック信号MNを

$$\text{MN1}[i][j] = \text{NN} / (\text{NY} + \text{NN}) \cdot \log((\text{MN1}[i][j] + \varepsilon) \cdot (\text{NY} + \text{NN}) / (\text{NN} \cdot (\text{MY1}[i][j] + \text{MN1}[i][j] + 2\varepsilon))) \quad \dots (11)$$

【0095】そして、更新された1次肯定メトリック信号MY1を1次肯定メトリック信号記憶部26に、更新された1次否定メトリック信号MN1を1次否定メトリック信号記憶部27に、新たに計算された肯定メトリック信号MYを肯定メトリック記憶部5へ、新たに計算された否定メトリック信号MNを否定メトリック記憶部6へ書き込む。以上で、KDメトリック学習部28は、メトリック学習の処理を終了し、メトリック学習制御信号MLCを学習制御部14に送る。学習制御部14は、KDメトリック学習部28からのメトリック学習制御信号MLCを受けて、学習制御部指示信号LIを1から0に変え、処理を終了する。

【0096】一度、以上の処理が行われると、辞書記憶部2の符号辞書が空でなくなるので、情報フィルタリングユニット50から出力される必要性信号N、信頼性信号Rは、0でなくなり、使用者の必要性の高い情報データが、未読データ記憶部10の上位に書き込まれるようになる。

【0097】以後、上記処理を繰り返すことにより、使用者が必要とする情報か否かを判定するために有効なキーワードが優先的に辞書記憶部2に記憶されるようになり、小規模な辞書であっても、精度の高い情報フィルタリングが可能となる。

【0098】なお、1次肯定メトリック信号MY1と1次否定メトリック信号MN1の学習を忘却の効果を入れた

$$MY1[i][j] = \alpha \cdot MY1[i][j] + LV[i] \cdot LV[j]$$

$$MN1[i][j] = \alpha \cdot MN1[i][j] + LV[i] \cdot LV[j]$$

を用いてもよい結果が得られる。（ここで、 $\alpha$ は、1より小さい正の数）もしくは、 $MY1[i][j]$ または $MN1[i][j]$ のいずれかが一定値をこれた場合に、

$$MY1[i][j] = MY1[i][j] / 2$$

$$MN1[i][j] = MN1[i][j] / 2$$

として、信号のオーバーフローを防ぐように構成することは、実施上好ましい。これは、適応符号辞書信号FDCK[j]の肯定回数PY[j]と否定回数PN[j]、および全肯定回数信号NYと全否定回数NNについても同様である。

【0099】さらに、文献「情報処理学会技術報告、自然言語処理101-8（1994.5.27）」などに記載された文書からキーワード群信号とキーワード数信号を生成するキーワード生成部を付加する構成をとれば、キーワードが与えられていない情報に対しても適用できる情報フィルタ装置を構成することができる。

【0100】また、タイトルがつけられた情報については、タイトルを構成する単語をもってキーワードとし、キーワード数信号とキーワード群信号を生成してもよ

い。加えて、キーワード信号は、国際特許分類番号など分類記号を含むようにもしても本発明の構成を変更する必要はなく、よい結果を得ることができる。

【0101】さらに、本実施の形態では、未読データURDを一つづつ提示する場合について示したが、表示装置（図示せず）の大きさによっては複数の未読データURDを同時に表示し、使用者がどの未読データURDについて応答したのかを正しく情報フィルタ装置に伝える構成をとることは容易である。

【0102】本発明の実施の形態2の情報フィルタの根幹は、キーワードの同時出現に注目したメトリックを導入することにより、キーワードという記号情報を距離の定義された空間に射影したことにある。これによって、キーワード群の遠近を距離というアナログ尺度で評価することができるようになり、これを利用することにより、従来の技術では必要か不要かの二者択一的な判定しかできなかつた必要性の評価が、ユーザーの必要性の順番に並べるといったことが可能になる。

【0103】加えて、本実施の形態2のように、ユーザーが必要とする情報の出現確率を用いて情報の必要性信号Nを計算する情報フィルタ装置において課題となっていた問題、すなわち、情報検索になれていないユーザーは、自分にとって必要な情報のみ「必要」と回答し、不要な情報については無視するために、ユーザーが必要とする情報の出現確率が1となってしまうという問題は、ユーザーが「必要」もしくは「不要」だけしか入力しない状態では、ユーザーが必要とする情報の出現確率NY／(NY+NN)を0もしくは1でない値、例えば0.5とすることにより解決することができる。

【0104】以上のように、本実施の形態による情報フィルタによれば、ユーザーからの教師信号に基づいた学習によって、ユーザーの必要とする情報に対しては、必要性信号が大きな値を取るようになり、その結果、表示装置等には、ユーザーにとって必要性が高い情報が優先的に表示されるようになる。

【0105】（実施の形態3）次に、本発明の第3の実施の形態について説明する。本実施の形態3は、実施の形態1または実施の形態2の構成における計算量およびメモリ量を、各メトリック信号を2次元の行列から1次元のベクトルとすることで、削減するものである。以下、本実施の形態3の動作を実施の形態1を参照して説明する。

【0106】本実施の形態では、肯定スコア信号SYは、

【0107】

【数4】

$$SY = \sum_{i=1}^{NWORD-1} V[i] \cdot MY[i] \cdot V[i]$$

... (12)

となり、否定スコア信号SNは、

【0108】

【数5】

$$SN = \sum_{i=1}^{nofDiC} V[i] \cdot MN[i] \cdot V[i]$$

... (13)

となる。また、教師データTD【c】の教師信号T【c】がT=1である場合には肯定メトリック信号MYの更新式は、

$$MY1[i] = MY1[i] + LV[i]$$

(ここで、i=1~nofDiC)

前記教師データTD【c】の教師信号T【c】がT=0である場合には、否定メトリック信号MNは、

$$MN1[i] = MN1[i] + LV[i]$$

(ここで、i, j=1~nofDiC)と更新される。その他は、実施の形態1と同様である。

【0109】なお、本実施の形態を採用する場合、辞書にキーワードの積(AND)などキーワード検索式含むように構成することにより、キーワードの共起をベクトル信号Vに反映させることは、非常に好ましい。また、この方法は、実施の形態2についても同様の方法で変更できる。

【0110】(実施の形態4) 次に、本発明の第4の実施の形態について図面を参照しながら説明する。本実施の形態4は、実施の形態1または実施の形態2または実施の形態3の情報フィルタ装置の構成にデータベース再構築装置部、データベース読み出し部及び適応データベース書き込み部等を附加し、情報フィルタ装置データベース再構築装置としたもので、実施の形態1のインタフェースユニット51、学習ユニット52及び情報フィルタリングユニットの機能を用いて使用者にとって必要な順にデータが並んだ使いやすい適応データベースを提供するものである。

【0111】図12に本実施の形態4のデータベース再構築装置プロック結線図を示し、以下に説明する。図12において、60はデータベース記憶部、61はデータベース記憶部60からデータを読み出し情報フィルタ装置に適した形にデータを整形して出力するデータベース読み出し部、62はデータベースの再構築を制御するデータベース再構築制御部、63はスイッチ、65は適応データベース記憶部、64は情報フィルタリングユニットからの信号を一時的に保持し最終結果を適応データ記憶部65に書き込む適応データベース書き込み部、200はデータベース再構築制御部62を制御する制御信号入力端子、201は学習データ数信号を入力する学習数信号入力端子である。その他のものは、実施の形態1に記載した情報フィルタ装置と同一の構成であるので省略する。

【0112】以上のように構成されたデータベース再構築装置の動作について説明する。まず、制御信号入力端子200からデータベース再構築開始を示す制御信号C

DB=1が入力され、学習数信号入力端子201から情報フィルタ装置の学習回数を示す学習数信号LNが入力される。データベース再構築制御部62は、データベース再構築制御指示信号出力端子210から出力されるデータベース再構築制御指示信号IRDを0から1に変え、処理中であることを示す。データベース再構築制御部62は、スイッチ63を情報フィルタリングユニット50と未読データ記憶部10とを接続するように切り替える。データベース再構築制御部62は、制御信号CD

10 B (=1) を受けて、データベース読み出し部61にデータベース記憶部60に記憶されたデータの数を聞く。データベース読み出し部61は、データベース記憶部60に記憶されたデータの数を数えその結果をデータ数信号nofDとして、データベース再構築制御部62に送る。データベース再構築制御部62は、適応データベース書き込み制御部64内の未読データ数nofURDの内容をデータ数信号nofDで置き換える。次にデータベース再構築制御部62は、学習数信号LNを読み出しデータ数信号RDNとして、データベース読み出し部61に送る。

20 【0113】データベース読み出し部61は、学習数信号LNを受けて、データベース記憶部60からLN個のデータを順次読み出し、必要な整形をして、情報フィルタリングユニット50に送る。情報フィルタリングユニット50は、実施の形態1に記載した動作を行い、未読データ記憶部10に格納する。

【0114】使用者は、インターフェースユニット51を起動し、未読データ記憶部10に格納されたLN個の未読データURDを順次読み出し、要不要を示す教師信号Tを入力する。LN個の未読データについての入力が終わると、使用者は学習開始信号入力端子106から学習開始信号LSを入力し、情報フィルタ装置の学習を行う。学習制御部指示信号出力端子107から出力される学習制御部指示信号LIが学習の終了を示すように1から0になると、データベース再構築制御部62は新たにLN個のデータを読み出すように読み出しデータ数信号RDNをデータベース読み出し制御部61に送り、新たにLN個のデータを情報フィルタリングユニット50を通し並べ替える。

40 【0115】使用者は、再び、インターフェースユニット51を起動し、LN個の未読データURDを必要か不要かを判断しながら、必要な情報が上位に来ているか否かを確認し、情報フィルタ装置にさらに学習させるか否かを決める。情報フィルタ装置の性能が不十分で、さらに学習させる場合には、使用者は、再び学習開始信号入力端子106から学習開始信号LSが入力し、情報フィルタ装置の学習を行う。

【0116】情報フィルタ装置の性能が十分に上がり学習が必要でなくなると、制御信号入力端子200からデータベース再構築実行を示す制御信号CDB=2を入力する。データベース再構築制御部62は、まずスイッチ

63を情報フィルタリングユニット50と適応データベース書き込み部64とが接続するように切り替える。次に、データベース再構築制御部62は、データベース記憶部60に記憶されたデータ数nofD個のデータを読みだすように読み出しデータ数信号RD\_Nをデータベース読み出し部61に送る。データベース読み出し部61は、nofD個のデータを順次読み出し、情報フィルタリングユニット50に送る。情報フィルタリングユニット50は、適応データベース書き込み部64の中のバッファにデータを必要性に基づいて並べ替える。

【0117】適応データベース書き込み部64は、書き込まれたデータの数がnofDになると、バッファの内容を適応データベース記憶部65に書き込み、書き込み終了信号EWをデータベース再構築制御部62に送る。書き込み終了信号EWを受けたデータベース再構築制御部62は、データベース再構築制御指示信号出力端子210から出力されるデータベース再構築制御指示信号IRDを1から0に変え、処理を終了する。

【0118】以上のように、本実施の形態4のデータベース再構築装置によれば、使用者にとって必要な順にデータが並んだ使いやすい適応データベースを作ることができる。

【0119】なお、本実施の形態4では、適応データベースは全体として元のデータベースと同じデータを持つようにしたが、記憶装置のメモリ領域を節約するために、適応データベースの内容をデータ間のリンク情報だけとしても、同じ効果が得られる。

#### 【0120】

【発明の効果】以上のように、本発明は、情報に割り振られた複数のキーワードをベクトルに変換する手段と、このベクトルと使用者からの教師信号を用いてスコアを計算する手段と、このスコアから必要性と信頼性を計算する手段とを備え、スコア計算部がスコアを計算する際に用いるメトリックを使用者から与えられる情報の必要/不要という単純な評価をもとに計算し、情報をユーザーの必要度にしたがって並べ、ユーザーに対して必要性の高い情報から順に提供することにより、初心者にも精度の高い情報を得ることができ、更に使用者にとって必要性の高い情報の取り出し易い情報フィルタ装置とデータベース再構築装置及び情報フィルタ方法と初期化方法を提供することができる。

#### 【図面の簡単な説明】

【図1】本発明の実施の形態1の情報フィルタ装置のブロック結線図

【図2】本発明の基本概念を説明するためのブロック結線図

【図3】本発明の実施の形態1の情報フィルタ装置のベクトル生成部の動作を説明するためのフロー図

【図4】本発明の実施の形態1の情報フィルタ装置の未読データ書き込み制御部の動作を説明するためのフロー

#### 図

【図5】本発明の実施の形態1の情報フィルタ装置の未読データ出力制御部の動作を説明するためのフロー図

【図6】本発明の実施の形態1の情報フィルタ装置の学習制御部の動作を説明するためのフロー図

【図7】本発明の実施の形態1の情報フィルタ装置のメトリック学習部の動作を説明するためのフロー図

【図8】本発明の実施の形態1の情報フィルタ装置の動作を説明するための模式図

【図9】本発明の実施の形態1の情報フィルタ装置の判定面学習部の動作を説明するための模式図

【図10】本発明の実施の形態2の情報フィルタ装置のブロック結線図

【図11】本発明の実施の形態2の情報フィルタ装置の辞書学習部の動作を説明するためのフロー図

【図12】本発明の実施の形態3のデータベース再構築装置の概略を示すブロック結線図

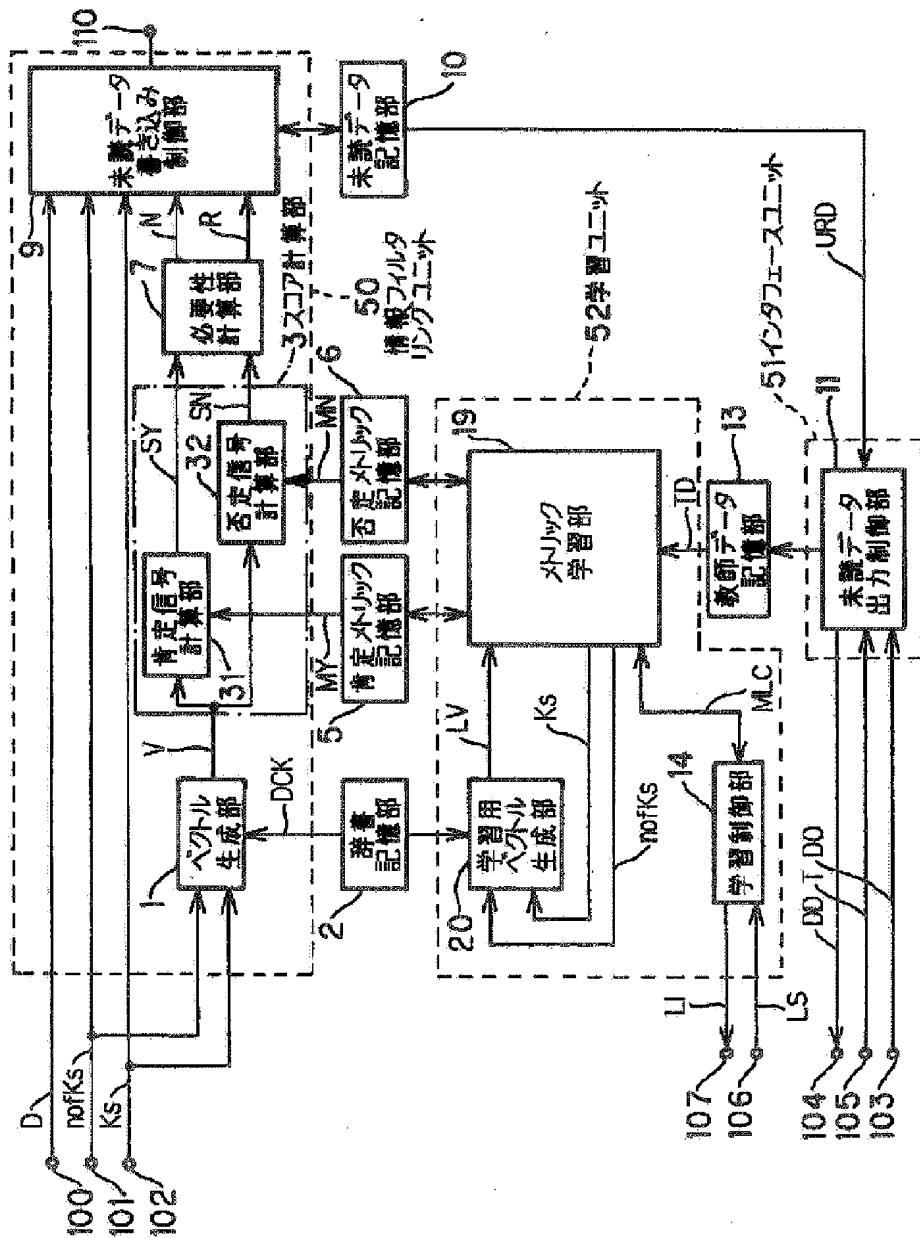
#### 【符号の説明】

- 1 ベクトル生成部
- 2 辞書記憶部
- 3 スコア計算部
- 5 肯定メトリック記憶部
- 6 否定メトリック記憶部
- 7 必要性計算部
- 9 未読データ書き込み制御部
- 10 未読データ記憶部
- 11 未読データ出力制御部
- 12 教師データ制御部
- 13 教師データ記憶部
- 14 学習制御部
- 19 メトリック学習部
- 20 学習用ベクトル生成部
- 23 辞書学習部
- 24 適応符号辞書記憶部
- 25 回数記憶部
- 26 1次肯定メトリック記憶部
- 27 1次否定メトリック記憶部
- 28 KDメトリック学習部
- 50 情報フィルタリングユニット
- 51 インタフェースユニット
- 52 学習ユニット
- 60 データベース記憶部
- 61 データベース読み出し部
- 62 データベース再構築制御部
- 63 スイッチ
- 64 適応データベース書き込み部
- 65 適応データベース記憶部
- 100 情報入力端子
- 101 キーワード数信号入力端子
- 102 キーワード信号入力端子

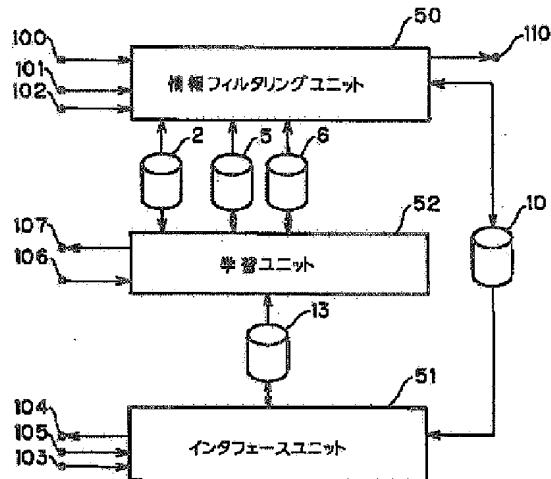
103 データ読み出し開始信号入力端子  
 104 データ表示端子  
 105 教師信号入力端子  
 106 学習開始信号入力端子  
 107 学習制御部指示信号出力端子

110 未読データ部指示端子  
 200 制御信号入力端子  
 201 学習数信号入力端子  
 210 データベース再構築制御指示信号出力端子

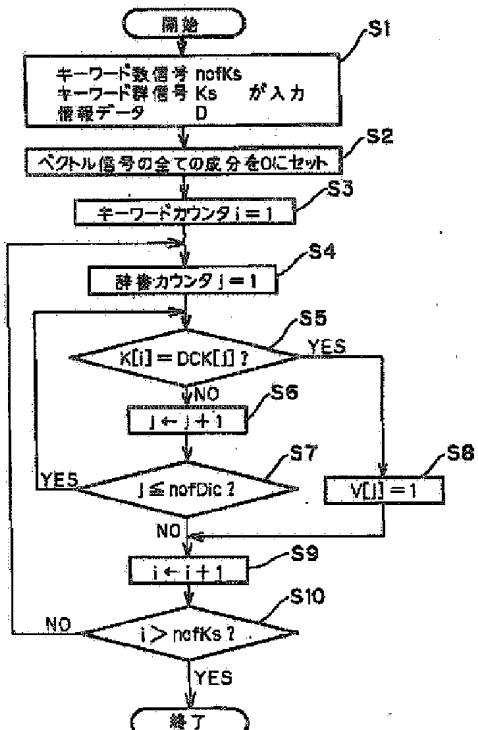
【図1】



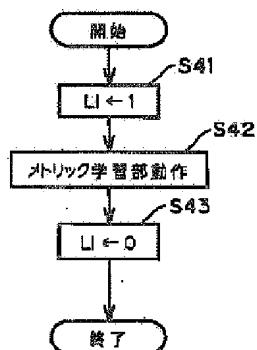
【図2】



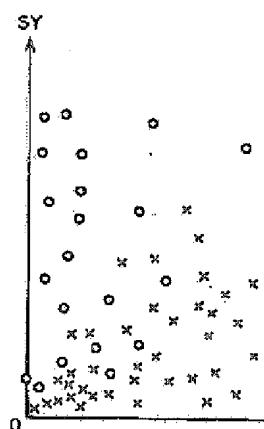
【図3】



【図6】

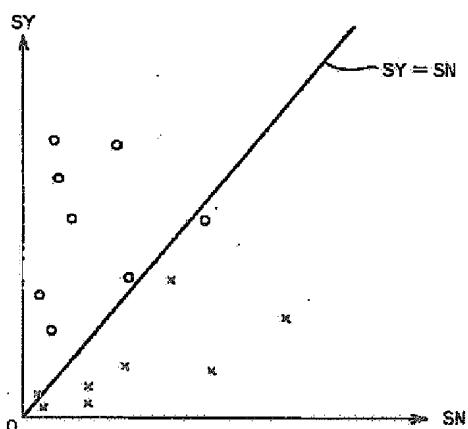


【図8】

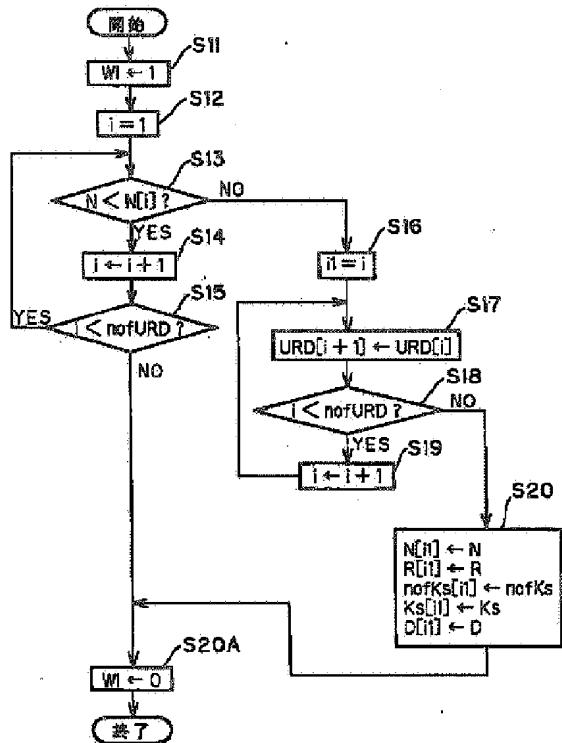


○  $T[C] = 1$   
 ×  $T[C] = 0$

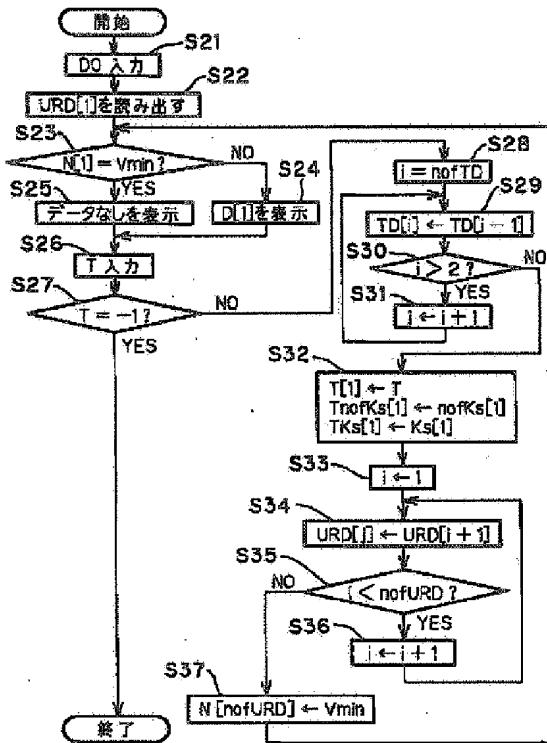
【図9】



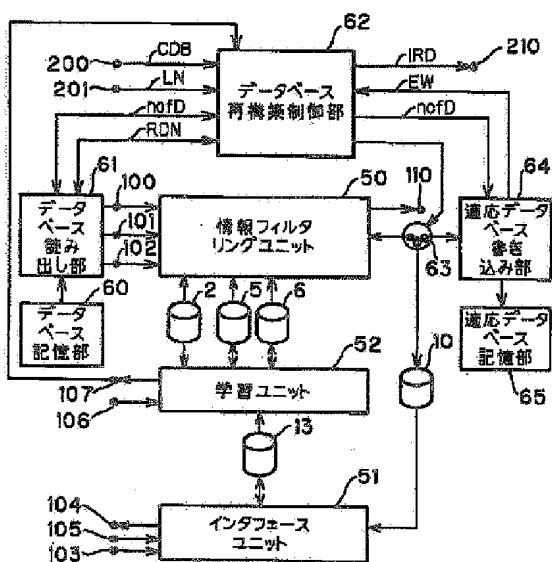
【図4】



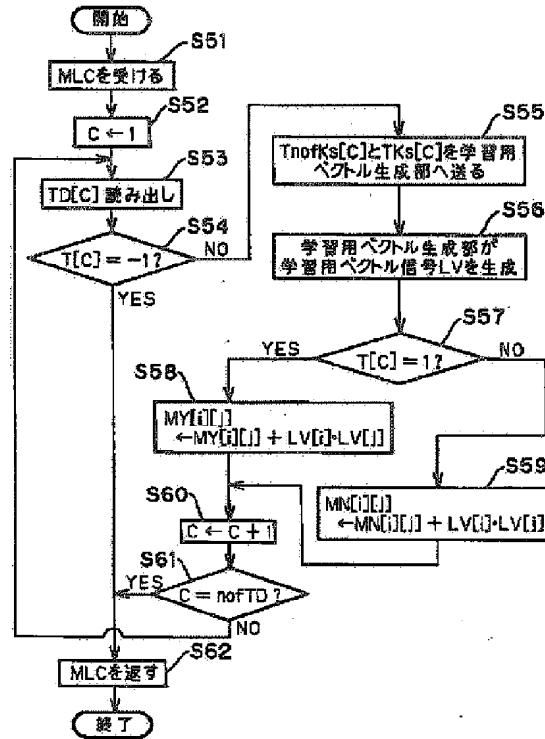
【図5】



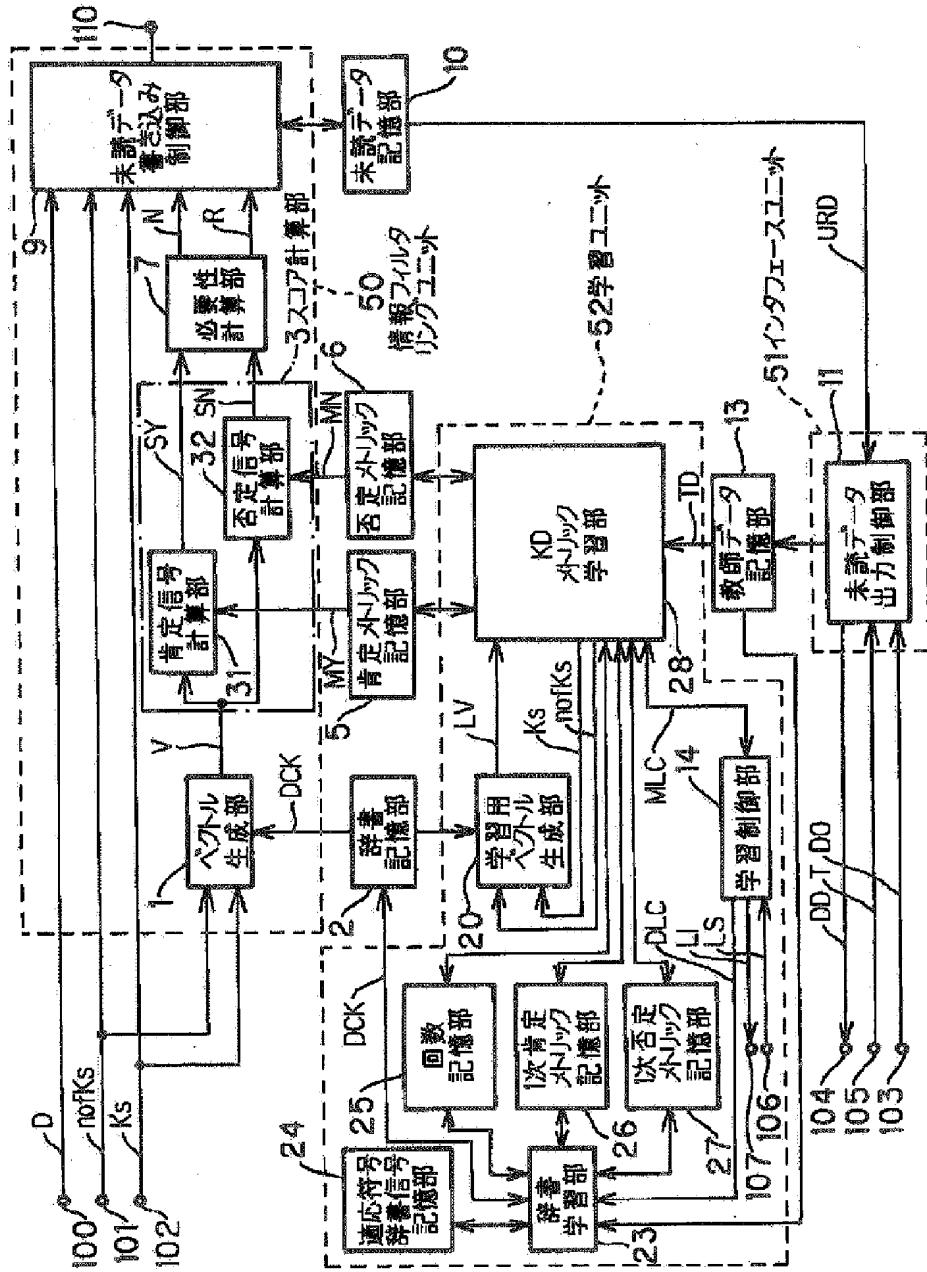
【図12】



【図7】



【図10】



【図11】

